An algorithm for dynamic routing in FMS under an unpredicted failure

Mahmoud A. Younis

Engineering Department, The American University in Cairo, Cairo, Egypt

Magdi S. Mahmoud

Department of Electronics and Communication Engineering, Cairo University, Giza, Egypt

The problem of dynamic part routing in an automated manufacturing system is described. The system consists of workstations capable of performing a number of different operations on a family of parts. In view of the unreliability of machines the production control system should manage the production schedule, given the random pattern of machine failure/repair. A three-level hierarchical control structure is suggested and simulated, using stochastic generation of events and part movements. The obtained results show that a production profile derived under the developed controller is feasible.

Keywords: flexible manufacturing systems, production control

Introduction

Modern manufacturing facilities vary more and more from traditional batch processing units, which have always had inherent limitations. The fundamental concept behind modern manufacturing is flexibility, which results from the constraints imposed upon the manufacturing environment by the market. The need for a short response time to change in demand, increasing variability in products, and a high production rate created the need for flexible manufacturing systems (FMS) as a concept. Integration of automated versatile machines, material handling systems, and advanced computer technology has led to the development of FMS for use in many different industries.[1-5] At present, FMS is capable of processing a large total volume of small to medium batches of parts.

The FMS performance is normally derived from simulation of appropriate models because actual experimentation on an existing system is not feasible. A comprehensive survey of the development of analytical models of FMS is presented in Ref. 1 and partially in Ref. 6. It has been pointed out that analytical models, sometimes referred to as generative models, are quite useful for systems with relatively few operating parameters. The effects of machine failures, demand uncertainties, and concurrent routing are difficult to cast. Other approaches include evaluative modelling,[7] perturbation analysis,[8] and network of queues.[9-12]

For modelling and simulation purposes, FMS can be regarded as a cluster of computer numerical control (CNC) machining centers, special machine tools, inspection machines, and some automated material handling devices, such as shuttle pallets, carts, and industrial robots.[13] The loading and unloading of machines, the adaptive route selection, part sequencing, and the determination of various manufacturing variables (such as feed rates and spindle speeds) are conducted under the real-time control of a host computer. A typical FMS can sequence parts randomly and respond quickly to the needs of the assembly floor because its setup time is minimal.[14] In Ref. 15 the production planning of FMS is divided into five problems, which can be solved sequentially or iteratively: part type selection, machine grouping, production ratio, resource allocation, and loading. Decision making in FMS is structured on three levels[16]: (a) long term—system configuration and hardware selection, parts-mix selection; (b) medium term—batching of parts and balancing of work in each batch; and (c) short term/real time—sequence of work into the system, sequence of operations, and reacting to failure and schedule changes.

Address reprint requests to Dr. Younis at the Engineering Dept., The American University in Cairo, P.O. Box 2511, Cairo, Egypt.

Received 10 January 1990; revised 4 November 1991; accepted 21 November 1991

Most manufacturing systems are large and complex. Therefore it is natural to divide the control or management into a hierarchical multilevel structure. Each level is distinguished by the length of the planning horizon and the kinds of data required for the decision-making process. Lower levels of the hierarchy typically have short horizons and use detailed information, whereas higher levels have longer horizons and use highly aggregated data. The nature of uncertainties at each level of control also varies. Planning and control activities usually involve managerial decision making that can be implemented into a three-level structure as displayed in *Figure 1*. Typically, the operational level is concerned with the short-time production profile, and hence it focuses on the joint determination of production rates of members of the part family while taking into account the effects of the demand, the level of the downstream buffer levels, and the reliability of workstations.

This paper examines the problem of dynamic part routing in FMS through computer simulation, using stochastic generation of events and part movements. Parts are loaded into the system such that the production goals are met. A three-level hierarchical control structure is suggested. A computational algorithm is described for a multiproduct-multiworkstation production system. A simulation example is presented to study the performance of this control policy.

The problem of dynamic part routing

Here we focus on the operational level and consider the flow control of determining the production rates for the part family. The horizon is set in relation to the master production plan.

Model development

The FMS consists of M workstations producing N different part types. Work station $m(m = 1, 2, \ldots, M)$ has L_m identical machines. Each part of type i requires K_i operations for its completion. A particular operation can be done at one or more different workstations. The time necessary to complete operation k on a type i part at workstation m is a random variable with mean τ_{im}^k. The flow rate of type i parts to station m for operation k is defined as $y_{im}^k(t)$.

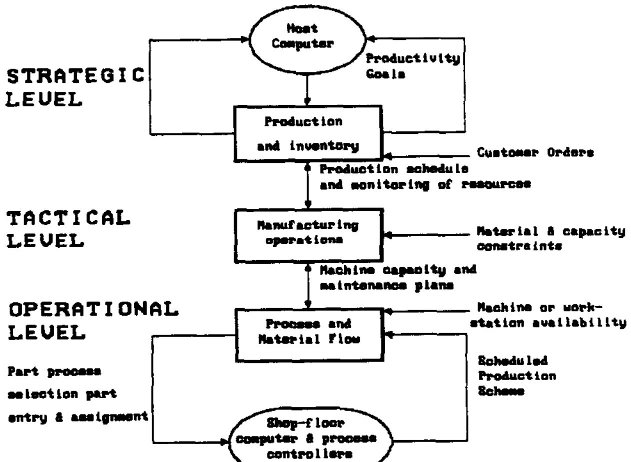

Figure 1. Three-level managerial decision making

Let u_i be the production rate of type i parts. Since no material is accumulated within the system, then conservation of flow implies that

$$\sum_{m=1}^{M} y_{im}^k(t) = u_i(t) \qquad k = 1, \ldots, K_i \qquad i = 1, \ldots, N \quad (1)$$

Define $\mathbf{x}(t) = [x_1(t) \ldots x_N(t)]$ as the buffer state that measures the cumulative difference between production and demand for the parts. Thus we have

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{u}_i(t) - \mathbf{d}(t) \qquad x_i(0) = x_{i0} \quad (2)$$

where $\mathbf{u}_i = [u_1(t) \ldots u_N(t)]$ is the vector of production rates for the part family and $\mathbf{d}_i = [d_1(t) \ldots d_N(t)]$ is the vector of downstream (terminal or intermediate) demand rates, which is known over the interval $(0,t_f)$, where t_f specifies the end of the production plan. The $\mathbf{d}(t)$ represents the production requirements set forth by the tactical level. Finished parts are stored in downstream buffers from which the downstream demand is satisfied. When $x_i(t) < 0$, backlogged demand occurs for part i, whereas $x_i(t) > 0$ gives the size of the inventory stored in the downstream buffers. The state of the workstations is called the machine state and is denoted by the integer vector $\boldsymbol{\alpha}(t) = [\alpha_1 \ldots \alpha_m(t)]$, the component $\alpha_m(t)$ representing the number of operational machines at station m. Recall that the production rate of any instant is limited by the capacity of the currently operational machines. We introduce $\Omega[\alpha(t)]$ to represent the capacity of the system and define it as the set of all production rates $\mathbf{u}(t)$ such that there exist feasible flow rates $y_{im}^k(t)$ satisfying (1) and

$$\sum_{i=1}^{N} \sum_{k=1}^{K_i} y_{im}^k \tau_{im}^k \leq \alpha_m \qquad m = 1, \ldots, M \quad (3)$$

The product $y_{im}^k \tau_{im}^k$ constitutes the proportion of each unit time interval used by one or more operational machines at station m to perform operation k on type i parts. From (3) the left-hand side is thus the total input work to station m per unit time due to the part flow rate y_{im}^k. The inequality reflects limited capacity.

Given that a machine at station m is operational, the probability of a failure in an interval δt is $(p_m\, \delta t)$. On the other hand, the probability that a failed machine is repaired during an interval δt is given by $(r_m\, \delta t)$, where p_m and r_m are the failure and repair rates, respectively, for the machines at station m.

Considering the structure of the FMS, an appropriate dynamic model of the machine state, where $P[E]$, represents the probability of occurrence of event E, is described by the following conditional probabilities:

$$P[\alpha_m(t+\delta t) = \sigma + 1 | \alpha_m(t) = \sigma] = \begin{cases} (L_m - 1)(r_m \delta t) \\ 0 \le \sigma < L_m \\ 0 \end{cases} \quad (4)$$

Otherwise

$$P[\alpha_m(t+\delta t) = \sigma - 1 | \alpha_m(t) = \sigma] = \begin{cases} \sigma(p_m \delta t) \\ 0 \le \sigma < L_m \\ 0 \end{cases} \quad (5)$$

It readily follows from (4) and (5) that

$$P[\alpha_m(t+\delta t) = \beta | \alpha_m(t) = \Delta] = 0$$

if $|\beta - \Delta| > 1$ and for two different machines states j and n, the quantity

$$\lambda_{jn}\delta t = P[\boldsymbol{\alpha}(t+\delta t) = n | \boldsymbol{\alpha}(t) = j]$$

defines the machine state transition probability which has the property that

$$\lambda_{jj} = -\sum_n \lambda_{jn}$$

The above discussion suggests modelling the times between failures (TBF) and the times to repair (TTR) by exponentially distributed random variables with mean time between failures (MTBF) and mean time between repairs (MTTR), where (MTBF $= 1/p_m$) and (MTTR $= 1/r_m$). Note from (4) and (5) that state transitions are due to the failure or repair of a single machine.

For convenience we consider that failure and repair rates are independent of production rates and the number of operational machines.

Problem statement

The dynamic part routing problem can now be stated. Given the buffer and machine states at arbitrary time, $\mathbf{x}(0)$ and $\boldsymbol{\alpha}(0)$, along with the dynamic model (1)–(5), we desire to determine the pattern of production over the interval $(0,t_f)$ such that the performance index,

$$J(\mathbf{x},\boldsymbol{\alpha},0) = E\left\{ \int_0^{t_f} g[\mathbf{x}(t)]\,dt \,\middle|\, \mathbf{x}(0) = \mathbf{x}_0 \quad \boldsymbol{\alpha}(0) = \boldsymbol{\alpha}_0 \right\} \quad (6)$$

is minimized where $E[,]$ is the expectation operator. The functional $g[\mathbf{x}(t)]$ penalizes the production profile for failing to meet the demand and for keeping an inventory of parts in the downstream buffers. In this regard, $J(\mathbf{x},\boldsymbol{\alpha},0)$ expresses the total penalty incurred by the producer in the interval $(0,t_f)$. For convenience we select $g[\mathbf{x}(\mathbf{t})]$ to be componentwise monotone and convex; that is,

$$\begin{aligned} g[\mathbf{x}(t)] &= \sum_j g_j[x_j(t)] \\ g_i(0) &= 0 \\ \lim g_i(x_i) &= \infty \qquad |x_i| \to \infty \\ g_{i-1}[x_{i-1}(t_{i-1})] &< g_i[x_i(t_i)] < g_{i+1}[x_{i+1}(t_{i+1})] \quad \text{for all } i \end{aligned} \quad (7)$$

Ideally, the optimal policy would be to produce at the demand rate, which means that the buffer state is kept at zero level. This ideal case is far from reality in view of the random failure of machines.

The optimal production policy

To characterize the optimal production policy, we consider the "cost to go" when the applied policy is $\mathbf{u}[\mathbf{x},\boldsymbol{\alpha},t]$ as

$$J^u(\mathbf{x},\boldsymbol{\alpha},t) = E\left\{ \int_t^{t_f} g[\mathbf{x}(s)]\,ds \,\middle|\, \mathbf{x}(t) = \mathbf{x}\,\boldsymbol{\alpha}(t) = \boldsymbol{\alpha} \right\} \quad (8)$$

Following the development of Rishel,[17] it can be shown that the optimal $\mathbf{u}^*$ is determined by

$$\min_{u \in \Omega(\boldsymbol{\alpha})} [dJ^u/d\mathbf{x}]\mathbf{u} = R \quad (9)$$

A systematic procedure to satisfy (9) subject to (1)–(7) results in a computational algorithm that will be described in the next section.

A computational algorithm

A general purpose computational algorithm has been developed for a multiproduct, multistation flexible manufacturing system. The algorithm is developed such that a randomly chosen step function is used to model the time-varying demand rate for each part type at a randomly generated interval within the time horizon of the production plan.

The times between failures and between repairs of the production resources are again modelled by exponentially distributed random variables. The means (MTBF $= 1/p_m$) and (MTTR $= 1/r_m$) are then computed from the randomly generated machine states $\alpha_m(t)$, where a reliability for each machine has been considered. The system is penalized equally for being ahead or behind demand requirements by a chosen cost function satisfying (7),

$$g[x(t)] = \sum_{i=1}^{N} \alpha_i x_i^2 \quad (10)$$

The buffer state $x_i(t)$ can be calculated by integration of (2); thus

$$x_i(t) = \int_t^{t_f} [u_i(t) - d_i(t)]\,dt \quad (11)$$

From (8) and (9) and by using (10) and (11) one can get

$$J^u(\mathbf{x},\boldsymbol{\alpha},t) = E\left\{\int_t^{t_f} \sum_{j=1}^{N} \alpha_j \int_0^s \{u_j(\tau) - d_j(\tau)\}^2 \, d\tau\right\} \begin{vmatrix} \mathbf{x}(t) = \mathbf{x} \\ \boldsymbol{\alpha}(t) = \boldsymbol{\alpha} \end{vmatrix} \quad (12)$$

and hence we obtain

$$R = \min_{u \in \Omega(\alpha)} [dJ^u/d\mathbf{x}]\mathbf{u} \quad (13)$$

It should be emphasized that algebraic manipulations of (11)–(13) are performed by numerical iterations.

A flow chart for the computational algorithm is given in *Figure 2*. Following is a brief description of the various routines in the computer program.

CONTROL. This is the main program that organizes the various computational steps of the algorithm. It also reads the title and all input data, such as time length of the production plan, workstation M, number of machines in each workstation, L_m, and number of part types, N, as well as the processing time for each part τ_{im}^k and time interval at which the flow control level implemented in the simulation is asked to compute the production data.

DEMAND. This generates the time-varying demand rate for each part by using a random number generator.

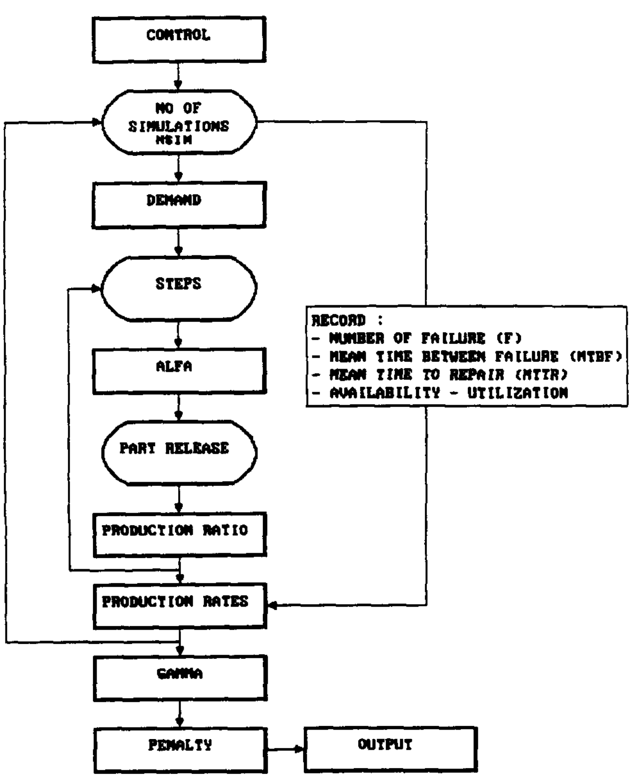

Figure 2. A flow chart for the computational algorithm

The computer code is based on

for part i to part N
TIME = RANDOM (1)
DEM = RANDOM (TIME)

STEPS. The production plan can be monitored by performing the calculations at the end of a previously determined time interval. The time horizon TF of the production plan is divided into F intervals. The computer code is of the form

TF = 300 h
ST = 30 min
Repeat for time interval = 0
till F = (TF * 60/ST) + 1

ALFA. All possible machine states are generated randomly, and the averages MTBF $= 1/p_m$ and MTTR $= 1/r_m$ are then evaluated.

PART RELEASE. This generates all possible combinations of parts entering the production line $u_s(1 \leq s \leq N)$ and $u_r(1 \leq r \leq N)$.

PRODRATIO. The production rate at each instant is limited by capacity of the currently operational machines. At time t the production rate must lie in a set $\Omega[\alpha(t)]$, which depends on the machine state. The program generates the production constraint sets for each part type combination at each machine state. The production vectors are then calculated.

PRODRATES. Calculate the production rate for each type u_i by assuming that parts production rates are simply proportional to the demand rates $(u_i d_i = u_j d_j)$ and solving with the previously obtained production vectors. The values obtained are then normalized and stored in an array.

GAMMA. A curve fitting is applied to the production rates calculated at each time interval, and a set of time function production rates for part types within the time horizon of the production plan is computed.

PENALTY. The system is penalized for backlogged demand or inventory of a part type by applying the cost function (10) and computing the result R from (12). The calculation is repeated NSIM times, and the minimum value of R is considered as an optimum value.

OUTPUT. This program organized the output results of the algorithm. It prints out the demand $d_i(t)$ and production rates $u_i(t)$ as well as the buffer state $x_i(t)$. The program gives the statistical data for the workstations. It computes the MTBF, the MTTR, and the availability and utilization of the available time at each workstation.

Simulation example

To demonstrate the application of the developed algorithm, we consider a typical case study: the manufacturing of windows, door, and curtain walls made out of aluminum profile. *Figure 3* shows the FMS plant, which consists of five workstations, as well as inspection and loading-unloading stations. The workstations include two profile cutting machines (A), three assembly machines (B), two machines for fastening and fitting the frames (C), two machines for mounting accessory fixtures (D), and two machines for glazing and sealing (E). A conveyer links the five working stations together with the inspection and loading stations. Special jigs and fixtures are used at stations A, B, and C in order to enhance the production process and to improve the quality. Seven part types are manufactured on this FMS. They are divided into three groups by making use of the group technology, production systems. The first group consists of three different sizes of windows and requires five operations—A, B, C, D and E. The second group consists of two sizes of doors and requires four operations—A, B, C, and E. The third group consists of two sizes of curtain walls and needs only three operations—A, B, and C.

The average processing times for each part type are given in *Table 1*. Initially, the system has all machines operating, and therefore the buffer state $x(t_0) = 0$. The demand ratio is taken to represent 80% of the system production capacity.

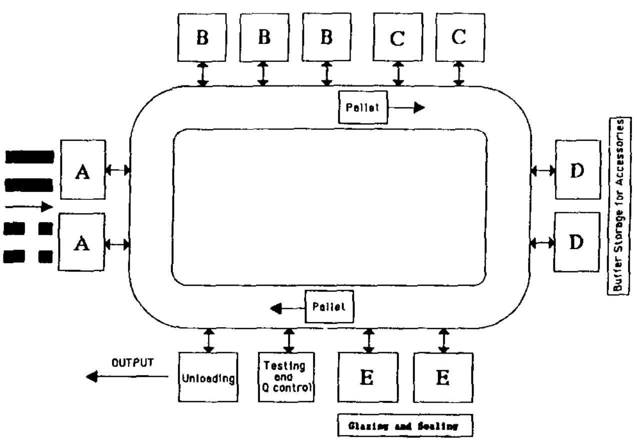

Figure 3. Flexible manufacturing plant. (A) Profile cutting; (B) assembly; (C) fastening and fitting; (D) mounting accessories; (E) glazing and sealing

Simulation results

The system of the plant given in *Figure 3* was considered in computer simulation of the three-level hierarchical control structure of *Figure 1* based on stochastic generation of events and randomized part movements. The computational algorithm of *Figure 2* managed the production schedule, considering the random pattern of machine failure/repair. The simulation model was run for an equivalent of 300 h. The flow control level implemented in the simulation computes the production rate $\mathbf{u}(t)$, the demand rate $\mathbf{d}(t)$, and the buffer state $\mathbf{x}(t)$ at 30-min intervals. It should be pointed out that the buffer state $\mathbf{x}(t)$ refers to the accumulated difference between the actual production level and the desired demand and can therefore take on negative values, indicating backlogs. Throughout the computer simulation the algorithm produces the following items for each workstation:

NOF = number of failures
MTBF = mean time between failures
MTTR = mean time to repair
ULT = average utilization time for full-load working
AVL = availability of the machines

Figures 4 and *5* give the production and demand rates as well as the actual buffer state for part types 1 and 5, respectively. Similar results are obtained for the other part types. Results show how the algorithm was able to control the production rate and let it track or follow the demand rate, thus keeping a small number of pieces in the downstream buffer. However, it can be seen from both figures that the accumulated buffer states oscillate around the zero level. The maximum value of the buffer state represents about ±4% of the production rate. Here the cost function penalizes the controller equally for excess production and for backlogged demand. It is usually preferred to keep the buffer close to zero when all machines operate with no failure and to clear the backlog (negative value of buffer state) in cases when failure occurs rather than to maintain

Table 1. Processing time for the parts in minutes

Part number	Part name	Working time (min) A	B	C	D	E	Total time
1	Window (size 1)	2.0	3.1	2.5	1.5	2.1	11.2
2	Window (size 2)	1.6	2.9	2.4	1.5	1.9	10.3
3	Window (size 3)	1.4	2.8	2.2	1.4	1.8	09.6
4	Door (size 1)	3.1	3.9	4.2	N.A.*	2.5	13.7
5	Door (size 2)	2.9	3.5	3.5	N.A.	2.3	12.2
6	Curtain wall (size 1)	2.2	4.5	3.5	N.A.	N.A.	10.2
7	Curtain wall (size 2)	2.0	4.1	2.5	N.A.	N.A.	08.6

A, profile cutting; B, assembly; C, fastening and fitting; D, mounting accessories; E, glazing and sealing.
* N.A. stands for not available.

high values of inventory to compensate for backlogs resulting from future failures. Such behavior can be improved by penalizing negative buffer states more than positive ones in the cost function.

In a series of simulation experiments the demand pattern is generated as a time function with random amplitude and operational period. Its pattern, however, is restricted to be a step function to reflect sudden changes in demand. The developed program is repeated several times till a consistent set of results is obtained. For simplicity we present the average behavior.

The statistical data for each workstation are given in *Table 2*. The simulated random number of failures for the machines within a workstation is given as F. It is readily seen that stations 4 and 5 are more reliable and have the fewest failures, 96 and 94, respectively. Consequently, both stations have the largest average mean time between failures (MTBF), 184.4 and 185.7 min, respectively. No significant difference can be observed between the mean time to repair (MTTR) and the availability for all five stations. However, the controller has been able to attain a utilization of about 82% for each of workstations 1, 2, and 3. On the other hand, stations 4 and 5 are lightly loaded with only 73% and 77% of the available time being used.

Conclusion

A problem of dynamic part routing in a flexible manufacturing system is described. The machines are unreliable. The production control system was able to meet random production requirements, while the machines fail and are repaired at random times. A three-level hierarchical control structure is suggested. A general computational algorithm is described to fit into existing factory management structure. The algorithm is structured to handle a multiproduct-multiwork station production system. The example for a plant manufacturing door, windows, and wall curtains made out of aluminum profile is described. The example and the results show that it is possible to accurately track demand requirements while maintaining a low inventory level. The control policy is applied to the whole FMS and not merely to the first machine that the part type encounters. This eliminates the buildup of parts inside the system, reducing the scheduling problems while considering repair and failure information.

Acknowledgment

The authors would like to thank the referees for their constructive comments and helpful suggestions.

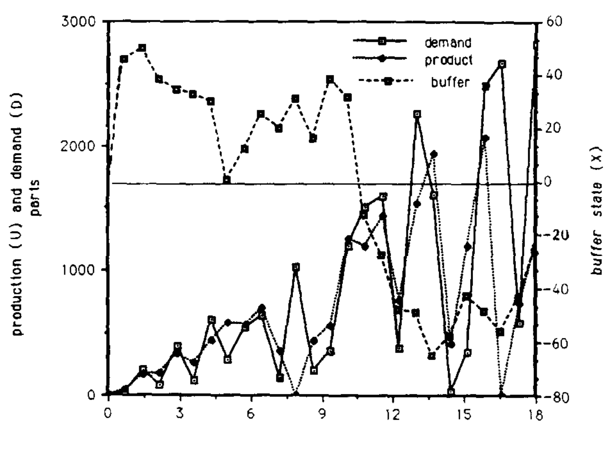

Figure 4. Production, demand and the actual buffer state

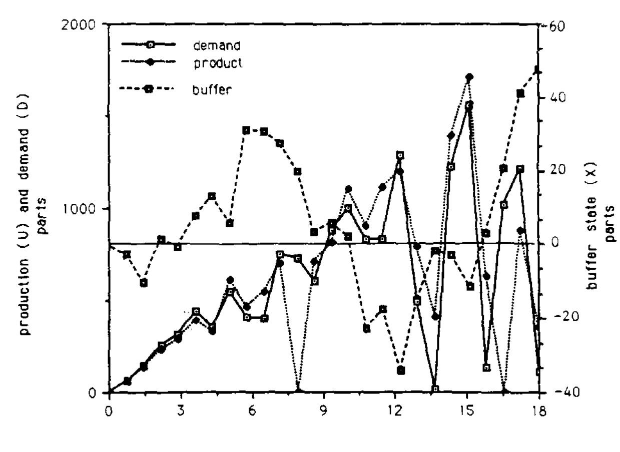

Figure 5. Production, demand and the actual buffer state

Table 2. Statistical data for workstations

Station	Number of failures, F	MTBF* (min)	MTTR** (min)	Availability (%)	Utilization (%)
1	123	145.4	30.2	83.7	82.6
2	114	155.3	30.3	89.3	82.5
3	100	177.0	33.3	85.2	82.7
4	96	184.4	34.1	85.3	73.3
5	94	185.7	32.9	86.3	77.9

* MTBF, mean time between failures.
** MTTR, mean time to repair.

References

1 Buzacott, J. A. and Yao, D. D. Flexible manufacturing systems: A review of analytical models. *Manage. Sci.* 1986, **32**, 890–905

2 Dupont-Gatelmand, C. A survey of flexible manufacturing systems. *J. Manuf. Syst.* 1982, **1**, 1–16

3 Klahorst, H. T. Flexible manufacturing systems: Combining elements to lower costs, add flexibility. *Indust. Eng.* 1981, **13**, 112–117

4 Suri, R. and Hildebrant, R. R. Modelling flexible manufacturing systems using mean-value analysis. *J. Manuf. Syst.* 1984, **3**, 27–38

5 J. Hatvany, ed. *World Survey on CAM*. Butterworths, Kent, UK, 1983

6 Buzacott, J. A. and Shanthikumar, J. G. Models for understanding flexible manufacturing systems. *AIIE Trans.* 1980, **12**, 339–350

7 Iwata, K., Murotsu, Y., Oba, F., and Yasuda, K. Production scheduling of flexible manufacturing systems. *CIRP Ann.* 1982, **31**, 319–333

8 Suri, R. and Dille, W. On line optimization of flexible manufacturing systems using perturbation analysis. *First ORSA/TIMS Conference on Flexible Manufacturing Systems*, Ann Arbor, MI, 1984

9 Bielecki, T. and Kumar, P. R. Optimality of zero-inventory policies for unreliable manufacturing systems. *Oper. Res.* 1981, **36**, 532–541

10 Gershwin, S. B. Representation and analysis of transfer lines with machines that have different processing rates. *Ann. Oper. Res.* 1987, **9**, 511–530

11 Kamath, M., Suri, R., and Sanders, J. L. Analytical performance models for closed-loop flexible assembly systems. *Int. J. Flexible Manuf. Syst.* 1988, **1**, 51–84

12 Suri, R. and Diehl, G. W. A variable buffer-size model and its use in analyzing closed queueing networks with blocking. *Manage. Sci.* 1986, **32**, 206–224

13 Groover, M. P. *Automation, Production Systems and Computer-Aided Manufacturing*. Prentice Hall, Englewood Cliffs, NJ, 1980

14 Arbel, A. and Seidmann, A. Performance evaluation of flexible manufacturing systems. *IEEE Trans. Syst. Man Cybern.* 1984, **SMC-14**, 132–140

15 Stecke, K. E. Formulation and solution of nonlinear integer production planning problems for flexible manufacturing systems. *Manage. Sci.* 1983, **29**, 273–288

16 Kimemia, J. and Gershwin, S. B. Flow optimization in flexible manufacturing systems. *Int. J. Prod. Res.* 1985, **23**, 81–96

17 Rishel, R. Dynamic programming and minimum principles for systems with Jumps Markov disturbances. *SIAM J. Control Optim.* 1987, **13**, 338–371